

DOCUMENT RESUME

ED 093 961

TM 003 779

AUTHOR Law, Alexander I.
TITLE Major Changes in the California State Assessment Program.
PUB DATE Apr 74
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April, 1974)
EDRS PRICE MF-\$0.75 HC-\$1.50 PLUS POSTAGE
DESCRIPTORS *Educational Assessment; Educational Legislation; Program Evaluation; *State Programs; State Surveys; Test Construction; *Testing
IDENTIFIERS *California State Assessment Program

ABSTRACT

According to a new law, state testing in California should be directed toward broad program evaluation rather than the diagnostic assessment of individual students which should be the responsibility of each local district. The data from the state testing program is used primarily for public information and to facilitate decision making at the state level. Four basic types of decisions are identified as needs assessment, funding decisions, funding exemplary programs, and program evaluation. The new legislation allows California to develop its own tests that can be made more relevant to California's needs than commercially available tests. Considerable effort, therefore, has been devoted to the specification of objectives that the test should assess. The steps involved in the process of test development are outlined. School means are the lowest level of analyses and multiple regression analysis was chosen to calculate expected scores from socio-economic and other background information. A number of developmental research projects will be conducted as the program is implemented. (RC)

MAJOR CHANGES IN THE CALIFORNIA STATE ASSESSMENT PROGRAM

By Dr. Alexander I. Law, Chief
Office of Program Evaluation and Research

California state testing has been changed by a new law which became effective in March, 1973. Although some aspects of the state assessment program remain unchanged (for example, testing will still be done in grades one, two, three, six and twelve), there are significant innovations of interest to educators and researchers.

Purpose of the State Testing Program

In the past, the state administered the tests to try to provide information for a wide range of audiences: state legislators, district administrators, program planners, classroom teachers, and the general public. In trying to meet the needs of such diverse audiences, ranging from the need of teachers for very specific diagnostic information about students to the more general needs for an indication of education's attainment statewide, the testing program did none of its jobs very well.

The stated intent of the new law is that state testing should be directed toward broad program evaluation rather than the diagnostic assessment of individual students. A state testing program can best be used to identify strengths and weaknesses of educational programs. It cannot meet the classroom need for individual diagnosis, which is the responsibility of each local district.

Purposes of State Testing

The purposes of state testing, identified by educators, are:

- A. to inform the public about how well children are learning basic skills and
- B. to facilitate decision making at the state level. Educators further identified four basic types of decisions:
 - 1. Needs assessment: to what extent are pupils of California and of each district mastering fundamental skills?
 - 2. Funding decisions: where are the greatest needs for extra resources and where will the allocation of extra help be most effective?
 - 3. Finding exemplary programs: which schools are attaining unusual success, and what factors appear to be responsible for that success?
 - 4. Program evaluation: are California pupils progressing significantly better because of the extra resources provided by programs such as Title I, Miller-Unruh, or Early Childhood Education?

Developmental Process for New Tests

One fundamental change under the new law is that California may develop its own tests rather than adopt a specific standardized test. This new testing program will involve the administration of a baseline test to grade 1; a reading test to grades 2 and 3, and basic skills tests to grades 6 and 12. The primary justification for spending time and resources at the State Department of Education for developing new tests is that they can be made more relevant to California's needs than commercially available tests. Considerable effort, therefore, has been devoted to the specification of objectives that the tests should assess. The steps in the process of test development are outlined below:

1. Assemble objectives

Although it would be possible to begin by writing objectives for each subject area, there would be considerable duplication of effort, and the objectives might not be appropriate for all schools in the state even if the authors were chosen from throughout California. It was therefore decided that objectives should be collected from the following sources: California state frameworks, textbook scope and sequence charts, commercial test publishers, and school districts in California. County and district superintendents were asked for copies of sets of objectives developed by their offices.

2. Combine objectives

A subject area specialist for each area was employed to aggregate the sets of objectives into one comprehensive list. The specialist needed to coalesce the diverse wordings of very similar objectives into a single statement of pupil performance.

3. Select relevant objectives

Statewide committees were formed to represent the following groups: school district curriculum specialists, teachers, offices of county superintendents, State Department of Education task forces, and professional associations and experts in the academic community. The committee selected those objectives which it felt were not important or relevant to the majority of California school districts. A comprehensive list of objectives were identified for each subject area.

4. Verify objectives

The final set of objectives for each subject area was sent to a random sample of districts to receive feedback for further improvement. All school districts in California were asked to respond in depth at some phase of this validation process.

5. Select test items

The State Department of Education is contracting with a number of test publishers to provide test items from their item pools matched to the set of objectives. The subject area advisory committees nominated teachers from various grade levels and geographical areas to serve on special item selection panels.

6. Subject test items to minority and linguistic critique

Items are then reviewed by representatives of ethnic and economic minorities to eliminate those items that appear to be culturally biased -- for example, reading passages in which vocabulary more familiar to one cultural group than another is used. Remaining items are then reviewed by linguists to detect items which use syntactic, structural or phonemic patterns which are unfamiliar to pupils from certain ethnic or language groups.

7. Field test

Before printing, the tests are field tested to ensure that directions to teachers and pupils are clear.

8. Develop a sampling plan

The law requires that every student in grades one, two, three, six, and twelve be tested. However, with the exception of the test for grade one, a matrix sampling procedure will be used. Matrix sampling was judged to be the most appropriate method of gathering a broad variety of information for program evaluation while minimizing the amount of testing time and associated costs. All students will be tested but only long enough to secure an adequate estimate of the performance of each school and district in the state, as required by statute. In grades two and three, for example, each pupil will take a 32 item test of which there are 10 unique forms. Each form measures all major aspects of reading, but focuses on different sub-skills.

The Entry Level Test for Grade One

One of the unique features of this comprehensive assessment program is the development of a baseline test for students in the first grade. The purpose of the test for grade one is to assess the skills children possess when they come to school.

The new test was developed by the staff of the Department of Education. It consists of the five subtests: Immediate Recall, Letter Recognition, Auditory Discrimination, Visual Description, and Language Development. The test was designed according to the recommendation of a legislative advisory committee that the test be a relatively short and easy one. Since the law forbids the use of individual pupil scores, they are not calculated or reported to schools. School means will be the lowest level of analyses.

This assessment will provide a basis for making judgments about the progress of schools and school districts on state achievement tests. In the past, the scores for each school and district have been compared with the state average regardless of initial differences in pupils readiness to learn or of differences in school resources for instructional programs. In the future, reports of test results will also reflect demographic characteristics such as poverty index, financial characteristics such as assessed valuation, and pupil characteristics such as socioeconomic level and pupil mobility.

The way this will be accomplished will be to predict the mean test score for a school or district from the information about the district and then to compare the observed score with the predicted or "expected" score.

Using Pupil and School Information to Calculate Expected Scores

The statistical method that is to be used to calculate expected scores is that of multiple regression analysis.

Regression analysis will derive the best set of weights for combining the socio-economic and other background information to predict school means. A band will be placed around the predicted score to help the reader avoid misinterpretations. Schools and districts will then be able to determine if their performance is above, below, or within the range of expected performance.

Special studies will be done to develop the most accurate prediction equations, for example, the use of moderator variables to form sub-groups of schools which have their unique equations or the use of special transformations of the data.

Reporting and Utilization

In addition to the use of prediction, other aspects of reporting bear mentioning. Since the test items will be drawn from a variety of published normative tests the results will be reported in terms of the average proportion of items answered correctly (P-value) for all items associated with each objective and sub-skill area. This value can easily be compared to the average P-value for the publishers' norm groups. It has the advantages of ease of understanding and can also serve as a type of "criterion" which can be used to show progress across years in absolute terms, while the predicted vs. observed index will show how a school compares to similar schools at a point in time.

No pupil-by-pupil information will be reported. For schools and districts, as much information will be reported as justifiable. For large schools and districts a profile of performance within each basic skill area will be reported for use by the district in program evaluation. At the state level, the maximum amount of information will be reported to program managers to assist in program revision, materials adoption, etc.

Associated Research Studies

A number of developmental research projects will be conducted as the program is implemented. At all grade levels,

several experimental versions of the test will be administered to a sample of pupils to provide a pool of items for periodic improvement of the tests. At the first grade level, studies now in progress are designed to indicate the constructive and predictive validity of the test, test-retest reliability, and any existing test or item bias. At the other grade levels, where matrix sampling is used, special studies are being made (1) to detect any content or sequence effects; (2) to develop special methods of assessing certain types of skills which ordinarily are measured by items requiring a common oral stimulus, e.g., word attack skills in reading; and (3) to determine the best method of computing an error estimate for a test which uses sampling with replacement in one part and sampling without replacement in the rest. Comparability tables will be developed at all grade levels for comparing scores on the state-developed tests and commonly used standardized tests.